THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

FIELD OF THE INVENTION

The present invention relates generally to a novel human gene and its derivatives and to mammalian, animal, insect, nematodes, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis and in conventional pharmaceutical preparations as well as in gene and protein replacement therapies.

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. Molecules of particular interest targeted by the inventors were gene regulators including regulatory proteins, signal transducters and heat shock proteins.

Gene expression generally requires interaction between a regulatory protein and an appropriate recognition sequence of a target gene. Regulatory proteins comprise in many cases a domain or motif which facilitates binding to DNA. One particular motif comprises small sequence units repeated in tandem with each unit folded about a zinc atom to form separate structural domains. This motif is now referred to as a zinc finger domain. Such a domain is generally defined by the number of cysteine (C) and histidine (H) residues.

In addition, knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction *via* receptors to intracellular transducers. One key signal transducer is Ras which couples the receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

Another regulatory mechanism involves heat shock proteins. The *Escherichia coli* heat shock protein, DnaJ, is the founding member of a family of proteins which are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG [SEQ ID NO:1] repeats and a C-terminal region with no obvious sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70 amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, a genes have been identified from the human genome which encodes proteins having a regulatory role. One gene, in accordance with the present invention encodes a protein with an N-terminal region resembling a zinc-finger domain of a novel type. Another gene encodes a protein involved in guanine nucleotide exchange factor (GEF) signalling pathways. Yet another gene encodes a protein which is a heat shock protein or heat shock-like protein which may have a role in tumour suppression.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Sequence identity numbers (SEQ ID NOs.) for nucleotide and amino acid sequences referred to in the subject specification are defined after the bibliography. A summary of SEQ ID NOs. is also given in Table 1.

One aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.

Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an $(HC_3)_2$ type.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:2;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:2 defines the gene, *mcg4*. This gene encodes a product, MCG4, having an amino acid sequence set forth in SEQ ID NO:3.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

A further aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:4 or 6 defines the gene, *mcg7*. This gene encodes a product, MCG7, having an amino acid sequence set forth in SEQ ID NO:5 or 7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an MCG7 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock binding protein or a derivative thereof.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:8;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 41°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:8 defines the gene, *mcg18*. This gene encodes a product, MCG18, having an amino acid sequence set forth in SEQ ID NO:7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG18 or a

derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

5

A summary of SEQ ID Nos. referred to in the subject specification is shown in Table 1.

TABLE 1

SUMMARY OF SEQ ID Nos.

SEQ ID NO.	DESCRIPTION
1	amino acid repeat sequence in DnaJ homologues
2	Nucleotide sequence of *mcg4*
3	amino acid sequence of MCG4
4	nucleotide sequence of *mcg7*
5	amino acid sequence of MCG7
6	nucleotide sequence of *mcg7* within exon of nucleotides 183-288
7	amino acid sequence of MCG7 within exon of nucleotide 183-288
8	nucleotide sequence of *mcg18*
9	amino acid sequence of MCG18
10-18	amino acid sequence identified using BESTFIT
19	sequence of pGEX and *mcg7* junction
20	sequence of pGEX and *mcg7* junction
21	nucleotide sequence of *myc*-tag/*mcg7* junction
22	amino acid sequence corresponding to SEQ ID NO:21
23	nucleotide sequence of pGEX and *mcg7* junction
24	amino acid sequence corresponding to SEQ ID NO:23
25-36	*mcg7*-specific oligonucleotide
37-45	*mcg18*-specific oligonucleotide

Single and three letter abbreviations for amino acid residues are shown in Table 2.

TABLE 2

Amino Acid	Three-letter Abbreviation	One-letter Symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Any residue	Xaa	X

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence [SEQ ID NO:2] and corresponding amino acid sequence [SEQ ID NO:3] of *mcg4*.

Figure 2 is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial murine expressed sequence tag (EST).

Figure 3 is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial nematode EST.

Figure 4 is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represent zinc atoms.

Figure 5 is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein.

Figure 6 is a representation showing that a related cysteine containing motif is present in the GATA-binding transcription factor from *Saccharomyces pombe*.

Figure 7 is a Northern blot showing expression of *mcg4* in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15μg total RNA derived from various human cancer cell lines. Lanes 1-5, respectively, contain RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

Figure 8 is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788.

Figure 9 is a representation of the partial nucleotide sequence alignment between a human

(W32939) and mouse (AA242159) *mcg4*-like EST in the putative 5' UTR of the *mcg4* cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

Figure 10 is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

Figure 11 is a diagrammatic representation of the domains of MCG4

zinc finger consensus: $CX_2HX_4CX_2CX_4HX_2CX_{17}CX_2CX_{18}HX_2CX_{18}CX_2C$

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged

basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged

leucine zipper domain consensus: $LX_6LX_6RX_6LX_6L$

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa261) $LX_6LXLX_6LXLX_6L$ (aa 286).

Figure 12 is a representation showing similarity of MCG7 with GEFs of various organisms.

Figure 13(a) is a representation of the nucleotide sequence [SEQ ID NO:4] and corresponding amino acid sequence [SEQ ID NO:5] of *mcg7*. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

Figure 13(b) is a representation of the partial nucleotide sequence [SEQ ID NO:6] and corresponding amino acid sequence [SEQ ID NO:7] of *mcg7* but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon of nucleotides 183-288.

Figure 14 is a representation showing a comparison between MCG7 and a homologue from *Caenorhabditis elegans* using the BESTFIT algorithm. In the figure, the following sequences are underlined:

EF-Hand= PROSITE DATABASE NO. PD0C00018

1a nematode DVDEEDEVEDIEF [SEQ ID NO:10]

1b human DVDGDGHISQEEF [SEQ ID NO:11]

nematode DHDRDGFISQEEF [SEQ ID NO:12]

1c human DQNQDGCISREEM [SEQ ID NO:13]

nematode DVDMDGQISKDEL [SEQ ID NO:14]

GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B

2 human HFVHVAEKLLQLQNFNTLMAVVGGLSHSSISRLKETH [SEQ ID NO:15]

nematode KFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY

[SEQ ID NO:16]

DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC

[SEQ ID NO:17]

nematode HNFHETTFLTPTTCNHCNKLLWGILRQGFKCKDCGLAVHSCCKSNAVAEC

[SEQ ID NO:18]

Figure 15 is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine *mcg7* cDNA (GenBank Acc. No. W71787 and AA237373). The putative initiation codon is underlined. The murine sequence represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

Figure 16 is a representation of further 5' nucleotide and corresponding amino acid sequence for human *mcg7*. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Fig. 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.

Figure 17 is a graphical representation of a GDP release assay. □ Experiment #1 (mean of duplicates). ◊ Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols

of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

K_m = 2.1μM, V_{max} = 37pMol/6min/36pMol [Expt#1]

K_m = 1.5μM, V_{max} = 30.3pMol/6 min/36pMol [Expt#2]

Figure 18 depicts various recombinant plasmids containing partial or full-length *mcg7*.

Figure 19 is a representation of the nucleotide sequence [SEQ ID NO:8] and corresponding amino acid sequence [SEQ ID NO:9] of *mcg18*.

Figure 20 is a representation showing that MCG18 has partial homology to *E. coli* DnaJ.

Figure 21 is a representation showing that MCG18 has homology to two *Caenorhabitis elegans* proteins.

Figure 22 is a representation showing that MCG18 has homology to a *Saccharomyces pombe* protein.

Figure 23 is a representation showing homology of MCG18 to a *Drosophila virilis* protein.

Figure 24 is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

Figure 25 is a representation of the nucleotide and corresponding amino acid sequence of murine *mcg18*.

Figure 26 is a representation of homology between human and murine MCG18.

Figure 27 depicts nucleotide sequences corresponding to the 5′ untranslated region of human *mcg18*.

Figure 28 depicts a Northern blot showing expression of *mcg18* transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain 15μg RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

5

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an $(HC_3)_2$ type.

Still more particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:2;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The present invention also provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

Another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock-binding protein or a derivative thereof.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:8;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least

about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

The nucleic acid molecule of the present invention defined by SEQ ID NO:2 is hereinafter referred to as constituting the "*mcg4*" gene. The protein encoded by *mcg4* is referred to herein as "MCG4"and has an amino acid sequence set forth in SEQ ID NO:3. The *mcg4* gene is proposed to encode, in accordance with the present invention, a regulator of gene expression and comprises a novel zinc finger domain, $(HC_3)_2$. A regulator of gene expression includes a transcription factor. Regulation may be at the level of nucleic acid:protein or protein:protein interaction.

The nucleic acid molecule of the present invention defined by SEQ ID NO:4 or 6 is hereinafter referred to as constituting the "*mcg7*" gene. The protein encoded by *mcg7* is referred to herein as "MCG7" and has an amino acid sequence set forth in SEQ ID NO:5 or 7 and is involved in signal transduction. The difference in the nucleotide and amino acid sequence is due to the presence or absence of an exon at nucleotides 183-288.

The nucleic acid molecule of the present invention defined by SEQ ID NO:8 is hereinafter referred to as constituting the "*mcg18*" gene. The protein encoded by *mcg18* is referred to herein as "MCG18" and comprises the amino acid set forth in SEQ ID NO:9.

The present invention extends to the naturally occurring genomic *mcg4, mcg7* and *mcg18* nucleotide sequences or corresponding cDNA sequences or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG4, MCG7 or MCG8 or the corresponding genetic sequences. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG4, MCG7 or MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to *mcg4, mcg7* or *mcg18*. "Additions" to the amino acid or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG4" or "*mcg4*", "MCG7" or "*mcg7*" or "MCG8" or *mcg18*" includes reference to all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG4, MCG7 or MCG18.

The *mcg4, mcg7* and *mcg18* of the present invention are particularly exemplified herein from humans and in particular from human chromosome 11q13.

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), reptiles, birds (eg. chickens, ducks, geese, parrots), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg. deer, foxes, kangaroos). Reference herein to *mcg4* and *mcg18* or their respective proteins MCG4, MCG7 and MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg4* gene portion of the genetic construct is operably linked to a promoter in the vector such that said promoter is capable of directing expression of said *mcg4* gene portion in an appropriate cell.

In addition, the *mcg4* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG4 is a transcription factor involved in gene regulation. Mutations in *mcg4* may result in aberrations in gene regulation leading to the development of or a propensity to develop various types of cancer. In this regard, although not wishing to limit the present invention to any one hypothesis or mode of action, it is proposed that *mcg4* or its expression product may be involved in the tissue-specific or temporal regulation of particular genes.

A deletion or aberration in the *mcg4* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may

be determined by assaying for aberrations in the parents and/or proband of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an *mcg7* polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg7* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg7* gene portion in an appropriate cell.

In addition, the *mcg7* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in *mcg7* or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg7* gene may also be important in the detection of cancer or

a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Yet another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg18* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg18* gene portion in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG18 is a transcription factor involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in tumour suppression. Thus mutations in *mcg18*

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in *mcg4, mcg7* or *mcg18* said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4, MCG7 or MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

A particularly convenient means of detecting a mutation in MCG4, MCG7 or MCG18 is by use of antibodies.

Accordingly another aspect of the present invention is directed to antibodies to MCG4, MCG7 or MCG18 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG4, MCG7 or MCG18 or may be specifically raised to MCG4, MCG7 or MCG18 or derivatives thereof. In the case of the latter, MCG4, MCG7 or MCG18 or their derivatives may first need to be associated with a carrier molecule. The antibodies to MCG4, MCG7 or MCG18 of the present invention are particularly useful as diagnostic agents.

For example, antibodies to MCG4, MCG7 or MCG18 and their derivatives can be used to screen for wild-type MCG4, MCG7 or MCG18 or for mutated MCG4, MCG7 or MCG18 molecules. The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG4, MCG7 or MCG18 levels or the presence of wild-type MCG4, MCG7 or MCG18 may be important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring certain therapeutic protocols.

As stated above antibodies to MCG4, MCG7 or MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

For example, specific antibodies can be used to screen for wild-type MCG4, MCG7 or MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG4, MCG7 or MCG18 in a cell extract or other biological fluid or purifying MCG4, MCG7 or MCG18 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4, MCG7 or MCG18 or to a specific mutant phenotype or to a deleted or otherwise altered region.

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG4, MCG7 or MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird with an effective amount of MCG4, MCG7 or MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG4, MCG7 or MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4, MCG7 or MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4, MCG7 or MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG4, MCG7 or MCG18 may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG4, MCG7 or MCG18 including cell extract

or tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG4, MCG7 or MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 37°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide

containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

As stated above, the present invention extends to genetic constructs capable of encoding MCG4,

MCG7 or MCG18 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which *mcg4*, *mcg7* or *mcg18* is involved in tissue-specific or temporal regulation.

Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and *mcg4*, *mcg7* or *mcg18* or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.

As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.

Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.

The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of *mcg18* expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier may be a solvent medium containing, for example, water, ethanol, polyol (for example,

glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as licithin and by the use of superfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimersal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 μg and 2000 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium

phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams, lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the

treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 μg to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 μg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

Effective amounts contemplated by the present invention include those amounts effective to ameliorate a condition. For example, it is envisaged that effective amounts would range from about 0.001 μg/kg body weight to about 100 mg/kg body weight. Alternatively, effective amounts of about 0.01 μg/kg body weight to about 10 mg/kg body weight or even 0.1 μg/kg body weight to about 1 mg/kg body weight. Administration may be per minute, hour, day, week, month or year or may only be a once off administration.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating *mcg18* expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18. Derivatives include fragments, parts, portions, mutants, homologues and analogues of the MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18. "Additions" to amino acid sequences or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18 immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 3.

TABLE 3

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva

	D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
	D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
5	D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
5	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
10	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
	L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L-α-methylalanine	Mala
15	L-α-methylarginine	Marg	L-α-methylasparagine	Masn
	L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
	L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
	L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
	L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
20	L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
	L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
	L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
	L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
25	L-α-methylserine	Mser	L-α-methylthreonine	Mthr
	L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr

L-α-methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl-ethylamino)cyclopropane	Nmbc		

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α-methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

Such analogues also apply in respect of MCG4 and MCG7.

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18. Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of MCG18 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression

molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression of MCG18 in a human, said method comprising contacting the *mcg18* gene encoding MCG18 with an effective amount of a modulator of *mcg18* expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of *mcg18*. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting Examples.

EXAMPLE 1

A human gene (designated *mcg4*) was identified on chromosome 11q13 that on the basis of sequence homology is predicted to encode a putative transcription factor of 310 amino acids (Fig. 1). *mcg4* is transcribed in several different cell lines (Fig. 7).

EXAMPLE 2

The expressed sequence tag (EST) database contains partial sequence data for the murine (Fig. 2) and nematode (Fig. 3) homologues of *mcg4*.

EXAMPLE 3

MCG4 contains a sequence of cysteine residues within the N-terminal region of the protein that resembles zinc-finger binding domains of a novel type, ie. $(HC_3)_2$ [Fig. 4].

EXAMPLE 4

Sensitive sequence homology searches reveal that related cysteine-containing motifs are present in another *C. elegans* protein (Fig. 5) as well as the GATA-binding transcription factor from *S. pombe* (Fig. 6).

EXAMPLE 5

mcg4 will have commercial value due to its likelihood of encoding a novel transcription factor that is highly conserved amongst organisms, thus suggesting an integral role in gene regulation. *mcg4* may also be involved in some way in tissue-specific or temporal regulation of certain genes, thus making it a potential target for modulating expression of those downstream effectors.

EXAMPLE 6

Nucleotide sequence data generated from cosmid clone cSRL-72c4 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) was aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al* 1990) and was found to match numerous human and mouse entries (Table 4 and Figure 2). These matching ESTs were further used to identify overlapping entries in the EST database (Table 5). The nucleotide sequences of these human ESTs were complied using MacVector 4.2.1 software (IBI-Kodak) to produce the cDNA sequence shown in Figure 1. EST entries AA074703 and AA134788 are closely related at the nucleotide level to *mcg4* and it is, therefore, likely that *mcg4* is a member of a newly discovered gene family (Figure 8).

The cDNA sequence of *mcg4* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) at the National Center for Biotechnology Information (http//www.ncbi.nih.gov.nlm). As the protein appeared to be novel, a translation of the longest reading frame for the *mcg4* cDNA was aligned to the EST database using the program TBLASTN, which performed a dynamic translation of the EST database in all 6 frames. The search results indicated that the nematode *C. elegans* had an MCG4-like protein (Figure 3), with the matching domains containing a spatial sequence of Cysteine and Histidine residues which resembled a zinc-finger structure (Figure 4). The program BLASTP was used, therefore, to conduct sensitive searches of the protein databases for similar zinc-finger motifs. A weak match to the putative zinc-finger domain was observed for another protein from *C. elegans* (Figure 5) and a poorer match for the GATA-binding transcription factor from *S. pombe* (Figure 6). The putative initiation codon of human *mcg4* is not preceded by an in-frame stop codon and it is therefore possible that the cDNA described in Figure 1 is a truncated form. However, sequence alignment of human and mouse *mcg4* ESTs showed a lower degree of nucleotide conservation prior to the assigned initiation codon, thus supporting the notion that the region represents the 5′ UTR (Figure 9). To determine the expression pattern of *mcg4*, 15μg of the total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% w/v MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer

using 20 x SSC (Sambrook *et al*, 1989). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (^{32}P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg4*. After washes in 0.1 x SSC/0.1% w/v SDS at 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg4* is expressed as a 1.6kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 7).

EXAMPLE 7

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 12).

EXAMPLE 8

The composite *mcg7* cDNA sequence is at least 2.4kb in length and Figure 13(a) shows a predicted translation product of at least 609 amino acids beginning at methionine 120. An alternative start site due to alternate exon splicing (indicated in lower case) may yield a protein of 671 amino acids starting at methionine 58 (Fig. 13a).

EXAMPLE 9

An *mcg7* homologue from *C. elegans* has been identified, the product of which is highly conserved with that of MCG7 (Fig. 14). There are several salient features of the protein which have been underlined in Fig. 14 - namely: a guanine nucleotide binding region, a diacylglycerol binding region, and "EF-hand"-calcium binding regions. In addition, there are several potential cAMP, protein kinase C, and casein kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

EXAMPLE 10

A number of partial human and murine EST clones exist for *mcg7*. The GenBank database

contains a cDNA (Acc. no. Y12336) encoding a full-length open reading frame (ORF) for human *mcg7* as well as a partial murine *mcg7* ORF (Y12339). In addition, the complete genomic sequence of the human *mcg7* gene is contained within GenBank entry AC000134.

EXAMPLE 11

The best characterised GEFs are members of the family of *ras* oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the *ras* signalling pathways. There is potential, therefore that the product of *mcg7* could also be a target for such clinical strategies.

EXAMPLE 12

The nucleotide sequence for *mcg7* cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683) (Fig. 16). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present (also shown in Fig. 13(a)). This closely matches the Kozak consensus. When this exon is absent, then the ATG is not in-frame and other possible initiation codons are absent (resulting translation shown in lower case lettering) (also shown in Fig. 13(b)). Further evidence that the initiation codon at position nt 681-683 is the true initiation site is given in Figure 15.

Alignment of human and a partial murine *mcg7* cDNA sequences is shown in Figure 15. The putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon and the sequence alignment thus suggests that this region represents the 5' UTR of *mcg7*.

Furthermore, similarity with the *C. elegans* homologue strongly suggest that the ATG codon at position nt 360-362 encodes the N-terminus of MCG7.

EXAMPLE 13

Figure 17 shows data from experiments indicating that a truncated version of MCG7 when expressed as a GST fusion protein (construct B in Fig. 18) can function as a Ras-guanine nucleotide exchange factor. In brief, Ras (unprocessed and as a GST fusion protein) is loaded with ^{3}H-GDP then incubated in the presence of excess cold GTP ± GST-MCG7. Full details of this assay can be found in Porfiri *et al.*

EXAMPLE 14

Nucleotide sequence data generated from cosmid clone cSRL-20h12 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) were aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al*, 1990) and was found to match GenBank entries T78563 (clone 113434) TO9103 (clone HIBBP12) and AA035643 (clone 471819). EST clones 113434 and 471819 were obtained from Genome Systems Inc. and these DNAs were sequenced on both strands with gene-specific primers (Table 5) to generate the cDNA sequence of *mcg7* shown in Figures 13(a) and (b).

The cDNA sequence of *mcg7* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) and the coding region was assigned on the basis of showing homology to the *C. elegans* protein F25B3.3 (Figure 14). The *mcg7* cDNA composite was suspected to contain a single nucleotide error that originated from clone 471819 and the correct nucleotide sequence was, therefore, sought by reverse transcription-polymerase chain reaction (RT-PCR) of the cDNA fragment from a human cDNA pool. Total RNA was extracted from a human lymphoblastoid cell line using an RNeasy Mini Kit (Qiagen). cDNA synthesis was conducted with the reverse transcriptase Superscript II RNaseH- (GIBCO, BRL) and random hexamers using the procedure recommended by the manufacturer (GIBCO, BRL). One fortieth of the cDNA mix was subjected to 35 cycles of PCR using the following cycling conditions: 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 90 seconds. The 50μl reaction mix consisted of 1x reaction buffer (Dade Scientific), 2mM dNTP mix, 20pmol of primers (see Table 6) MCG7UF (within the

variably spliced exon of Figure 13(b), between nucleotide positions 184-201) and SGCADRV2 (between nucleotide positions 866-846 of Figure 13(a)) and 10 units of Dynazyme (Dade Scientific). The resulting PCR product was cloned into the pGEM-T vector (Promega) using standard methodology and sequenced using gene-specific primers. The correct nucleotide sequence of *mcg*7 (as shown in Figure 13(a)) matches that of the recently release GenBank entry Y12336. A partial mouse *mcg*7 cDNA sequence can also be found in GenBank entry Y12339.

EXAMPLE 15

The coding sequence of *mcg*7 was cloned into vectors for expression in both bacterial and mammalian cells. In addition to the full-length constructs, the deletion constructs shown in Figure 18 were designed to retain the guanine nucleotide exchange (GEF) domain. For prokaryotic expression, the *mcg*7 coding region was inserted downstream of and in-frame with the Sj26 cassette of the pGEX (Pharmacia) series of vectors (Smith and Johnson, 1988) using standard cloning techniques (Sambrook *et al*, 1989). For mammalian expression, the *mcg*7 coding sequence was first *myc*-tagged at the N-terminus and then ligated into the expression vector pc Exv-n using standard cloning techniques. Ligation junctions of the constructs were sequences as the cloning strategies inadvertently changed or introduced additional amino acids as shown below.

Construct (A): EST clone 113434 was digested with *Apa*I (Figure 13(a), nucleotide positions 1022 to >2416 (within the vector)), blunt-ended with T4 DNA polymerase according to the specifications of the manufacturer (New England Biolab) and ligated into the *Sma*I site of pGEX-3X.

Sequence of the pGEX and *mcg*7 (underlined) junction:

pGEX-3X *mcg*7 (1022)

Sj26 ... GGG ATC CCC CTG GTC [SEQ ID NO:19]

additional amino acids Gly Ile Pro

Construct (B): EST clone 113434 was digested with *Eco*RI (Figure 13(a), nucleotide

positions <695 (within the vector) to 1711) and ligated into the *Eco*RI site of pGEX-1.

Sequence of the pGEX and *mcg*7 (underlined) junction:

pGEX-1 *mcg*7 (695)

Sj26 ... GAA TTC GGC ACG AGC CGA CGG [SEQ ID NO:20]

additional amino acids Glu Phe Gly Thr Ser

Construct (C): full-length *mcg*7: The pGEM-T clone containing the 5' end of the *mcg*7 coding region was digested with *Apa*I (subsequently blunt-ended with T4 DNA polymerase) and *Bst*XI to liberate the fragment between nucleotide positions 336 and 830 of Figure 13(a). Clone 113434 was digested with *Bst*XI and *Hind*III (vector derived) to liberate a fragment between nucleotide positions 830 > and 2416 (vector derived) of Figure 13(a). A pGEM-11zf vector (Promega) containing the *myc*-tag was digested with *Apa*I (subsequently blunt-ended with T4 DNA polymerase) and *Hind*III, and ligated with the 2 inserts described above.

Sequence of the *myc*-tag/*mcg*7 junction [SEQ ID NOs:21/22]:

```
----------myc-tag---------------     vector BamHI   mcg7 5' UTR (337)      start
ATGGAGCAGAAGCTGATCTCCGAGGAGGACCTG CCCGGGGCAGCTggatccG CAGCCCACCCCCGCGCCCGCGGCCATG
M  E  Q  K  L  I  S  E  E  D  L   P  G  A  A  G  S   A  A  H  P  A  P  A  A  M
                                 -------------additional amino acids---------
```

The *myc*-tagged full-length *mcg*7 insert in pGEM-11zf was then excised with *Sac*I and *Hind*III (both vector derived) and directionally cloned into the mammalian expression vector pEXV (Beranger *et al*, 1994).

Construct (D): Construct (C) in pGEM-11zf was sequentially digested with *Hind*III (this site was subsequently blunt-ended with T4 DNA polymerase) then *Bam*HI, and ligated into pGEX-2T digested with *Bam*HI and *Sma*I. Digestion with *Bam*HI, and ligated into pGEX-2T digested with *Bam*HI and *Sma*I. Digestion with *Bam*HI removed the *myc*-tag of Construct (C).

Sequence of the pGEX and *mcg*7 [SEQ ID NO:23/24] (underlined) junction:

```
pGEX-2     BamHI mcg7 (337)
Sj26  ...  gga tcc GCA GCC CAC CCC GCG CCG GCG GCC ATG
           Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met
           ----------additional amino acids----------
```

EXAMPLE 16

Overnight bacterial cultures containing the pGEX plasmid were used to inoculate 500ml of Luria Broth media containing 50μg/ml ampicillin. The cultures were grown to an OD of ~0.8 and then induced with 1mM of IPTG for up to 3 hours at 37°C. The bacteria were pelleted and resuspended in 15 ml of STE buffer (10mM Tris pH 8.0, 150 mM NaCl and 1mM EDTA) with 1 mg/ml lysozyme. The mixture was left on ice for more than 1 hour and subsequent steps were performed at 4°C. Protease inhibitors aprotinin, pepstatin and leupeptin were added at final concentrations of 25μg/ml, prior to the addition of Triton-X-100 (2% v/v final) and n-lauroyl sarcosine (1.5% w/v final). The lysate was sonicated for ~1 minute and pelleted at 14,000 x g for 15 minutes. 100 μl of 50% w/v glutathione-sephadex bead slurry (in PBS) was added per ml of supernatant. Following a 30 minute incubation at 4°C, the beads were washed three times with NETN (20mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5% NP40), once with NETN-HS (equivalent to NETN but with 1M NaCl), and once in NETN. The bound protein was directly analysed by SDS-polyacrylamide gel electrophoresis (PAGE) as described below or the bound protein was eluted from the beads with the following elution buffer (50mM Tris pH 8.0, 150mM NaCl, 5mM $MgCl_2$, 1mM DTT, 10mM reduced glutathione) for use in GDP release assays.

EXAMPLE 17

Twenty microlitres of GST-sepharose-bound MCG7 were added to an equal volume of 2 x sample loading dye (100mM Tris pH6.8, 2% v/v mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), boiled for 5 min and loaded onto a 7.5% w/v SDS-PAGE gel (Sambrook *et al*, 1989). The Coomassie brilliant blue stained gel (Sambrook *et al*, 1989)

typically displayed a protein doublet, running between 87-95 kDa consisting of the MCG7-GST fusion and a slightly smaller, co-purified contaminating *E. coli* protein of ~105kDa. The calculated molecular weight of full-length MCG7 is 77.5 kDa (Construct (D)) and the GST component has a molecular weight of 26kDa, hence, the recombinant protein runs slightly smaller than predicted. A Western blot of the same gel probed with anti-GST antibody yields an MCG7-specific band at the same position as that of the stained gel.

EXAMPLE 18

Assumptions: (a) GST-Ras molecular weight = 50 kD; (b) Concentration of GST-Ras solution = 1mg/ml = 20μM; (c) [^{3}H]-GDP is 1mCi/ml and 13.3Ci/mmol, therefore [H]-GDP concentration = 75 μM and 1pmol [^{3}H]-GDP=15,466 cpm; (d) Elution buffer = Buffer E = 20 mM Tris-Cl, pH7.5; 50mM NaCl; 5mM $MgCl_2$; 1mM DTT (added just before use). Buffer E + BSA= Buffer E+1mg/ml BSA (added just before use).

Mix together, in the following order and mix well after each addition:
10μl (=10μg) GST-Ras (@1mg/ml in Buffer E), 463μl Buffer E + BSA, 7μl [^{3}H]-GDP, 10ml 490 μM EDTA. Incubate @ RT for 10 min. Add 10μl 0.5 M $MgCl_2$ and mix well. Incubate @ RT for 10 min. Place on ice. During the first incubation the excess EDTA concentration is 5mM, during the second incubation the excess Mg concentration is 5mM. The [^{3}H]-GDP concentration is 1μM and the final concentration of GST-Ras is 400nM. Thus 20ml of the final mix will contain 8pmol of GST-Ras protein. Specific activity of GDP is 15,446 cpm/pmol x (1/1.4) = 11,047 cpm/pmol.

EXAMPLE 19

Exchange Ras with labelled GDP as above. Add unlabelled GTP (stock = 100mM, pH7) to 1 mM. Adjust Mg concentration by adding 5μl 0.5 EDTA to labelled Ras, 5μl 0.5M EDTA to 500μl MCG7, and 5μl 0.5M EDTA to 500μl Buffer E + BSA. On ice set up microfuge tubes with 40μl Ras-GDP (in triplicate) with 40μl MCG7 or Buffer E + BSA (control). Transfer tubes to heat block @ 25°C and incubate for 10, 20 or 30 min. Stop exchange reactions with 1ml of

ice cold buffer E and place on ice. Pre-soak nitrocellulose filters, pore size 45μm, in Buffer E. Assemble the vacuum manifold apparatus (Millipore) with wet filters and plug the wells with rubber bunds. Switch on the vacuum pump. Remove the first plug, aliquot the sample and once it has been sucked through, wash the filter with 10ml of ice cold Buffer E. Remove next plug etc and continue round the manifold. Take manifold apart. Pin the filters to a pin board reserved for [^{3}H]. Air dry. Take up in 4ml scintillation fluid and count. These studies have been carried out with a truncated MCG7-GST fusion protein (amino acids 341 of Figure 13a to stop encoded within construct B).

EXAMPLE 20

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (*mcg18)* is expressed as an ~1.4kb mRNA (Fig. 28) and is predicted to encode a 241 amino acid product (Fig. 19).

EXAMPLE 21

MCG18 has partial homology to *E. coli* dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 20).

EXAMPLE 22

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 21) and *S. pombe* (Fig. 22) that also feature the J domain but maintain sequence similarity through the central and C-terminal regions of the proteins.

EXAMPLE 23

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) 70 and consist of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 23) functions as a

tumour suppressor.

EXAMPLE 24

A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 24.

EXAMPLE 25

During the sequence characterisation of the *VRF/VEGFB* promoter region on cosmid CLGW4 [Grimmond *et al*, 1996], which maps to chromosome 11q13 the inventors identified a sequence that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated *mcg18*. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse *mcg18* (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in Table 7. The EST clones listed in Table 8 were also utilised in generating the full-length coding sequence for human (Figure 19) and mouse (Figure 25) *mcg18*. The EST database also contained *mcg18* cDNA entries that were alternately (or partially) spliced, and in order to understand their ability to encode new polypeptides, the gene structure of *mcg18* was determined by sequencing human and mouse genomic templates with gene-specific primers.

Genomic fragments containing the human [Grimmond *et al*, 1996] and murine genes [Townson *et al*, 1996] have been previously reported. Cosmid CLGW4 contains the entire human gene and λ121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 7. Plasmids containing sub-fragments of λ121 and cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond *et al*, 1996; Townson *et al*, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul *et al*, 1990] was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (http//www.ncbi.nih.gov.nlm). The sequence

data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

The cDNA sequence of human *mcg18* (Figure 19) was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul *et al*, 1990] and the coding region was identified on the basis of showing homology to the DnaJ family of proteins (Figure 20). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure 27). Similar database search results were obtained for the mouse *mcg18* cDNA, and the alignment of human and mouse protein sequences is shown in Figure 26. MCG18 has greatest homology to gene products from *C. elegans* (Figure 21) and *S. pombe* (Figure 22). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from *D. virilis* (Figure 23), nor is it a homologue of other reported human J-domain-containing proteins (Figure 24).

To determine the expression pattern of *mcg18*, 15μg of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook *et al*, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (^{32}P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg18*. After washes in 0.1 x SSC/0.1% w/v SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg18* is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 28).

TABLE 4

ESTs matching *mcg4*

accession number	seq. run	organism	score	E value	N
gb|AA399110|AA399110	zt89e06.s1	Soares testis NHT Homo sa...	1136	4.0e-168	2
gb|N39612|N39612	yy51g06.s1	Homo sapiens cDNA clone 2...	1521	5.3e-168	4
gb|AA514406|AA514406	nf57d01.s1	NCI_CGAP_Co3 Homo sapiens...	931	5.5e-166	3
gb|AA544946|AA544946	vk38e02.r1	Soares mouse mammary glan...	1207	8.4e-164	2
gb|AA450076|AA450076	zx42a04.s1	Soares total fetus Nb2HF8...	691	2.3e-160	4
gb|AA535731|AA535731	nf88f07.s1	NCI_CGAP_Co3 Homo sapiens...	796	3.5e-158	4
gb|W79710|W79710	zd86f01.r1	Soares fetal heart NbHH19...	1644	1.1e-157	4
gb|AA503531|AA503531	ne47e08.s1	NCI_CGAP_Co3 Homo sapiens...	736	4.0e-156	4
gb|AA450132|AA450132	zx42a04.r1	Soares total fetus Nb2HF8...	1955	3.9e-155	1
gb|AA398068|AA398068	zt89f06.r1	Soares testis NHT Homo sa...	1315	5.4e-148	2
gb|W60405|W60405	zd29h08.r1	Soares fetal heart NbHH19...	1022	1.8e-139	4
gb|W81382|W81382	zd86f01.s1	Soares fetal heart NbHH19...	605	3.5e-125	5
gb|AA047617|AA047617	zf13f07.s1	Soares fetal heart NbHH19...	922	4.6e-125	2
gb|AA282175|AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapien...	1577	2.0e-123	1
gb|AA242159|AA242159	my30d04.r1	Barstead mouse pooled org...	866	7.7e-117	2
gb|AA068680|AA068680	mm61a05.r1	Stratagene mouse embryoni...	1280	1.6e-98	1
gb|W46766|W46766	zc36b07.s1	Soares senescent fibrobla...	506	9.6e-92	3
gb|N93704|N93704	zb51c04.s1	Soares fetal lung NbHL19W...	584	9.0e-91	4
gb|AA155210|AA155210	mr98e01.r1	Stratagene mouse embryoni...	840	7.6e-87	2
gb|AA366022|AA366022	EST76915	Pineal gland II Homo sapien...	1077	2.4e-81	1
gb|AA037691|AA037691	zk34h12.s1	Soares pregnant uterus Nb...	949	2.1e-80	2
gb|W35374|W35374	zc07h03.s1	Soares parathyroid tumor ...	1016	3.1e-76	1
dbj|C00696|C00696	HUMGS0008251,	Human Gene Signature, ...	1009	1.2e-75	1
gb|T98249|T98249	ye59a07.s1	Homo sapiens cDNA clone 1...	998	6.7e-75	1
gb|W21588|W21588	zb51c04.r1	Soares fetal lung NbHL19W...	484	1.1e-69	4
gb|H32171|H32171	EST107015	Rattus sp. cDNA 5' end.	828	1.1e-60	1
gb|AA108092|AA108092	mm89e06.r1	Stratagene mouse embryoni...	782	1.3e-60	2
gb|AA017857|AA017857	mh44d10.r1	Soares mouse placenta 4Nb...	665	2.5e-60	2
gb|AA037690|AA037690	zk34h12.r1	Soares pregnant uterus Nb...	540	9.4e-53	2
gb|AA531006|AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapien...	535	5.4e-48	2
gb|N46760|N46760	yy51g06.r1	Homo sapiens cDNA clone 2...	665	9.5e-47	1
gb|W23584|W23584	zc71d03.s1	Soares fetal heart NbHH19...	457	1.8e-44	2
gb|W42214|W42214	mc69h09.r1	Soares mouse embryo NbME1...	460	1.3e-38	3
gb|AA244877|AA244877	mx25a04.r1	Soares mouse NML Mus musc...	429	2.9e-25	1
gb|W32939|W32939	zc07h03.r1	Soares parathyroid tumor ...	320	4.8e-18	1

TABLE 5

ESTs matching AA074703 (*mcg4*-related cDNA)

Database: Non-redundant Database of GenBank EST Division
1,222,625 sequences; 449,352,662 total letters.

Sequences producing High-scoring Segment Pairs:		High Score	Smallest Sum Probability P(N)	N
accession number	seq. run organism	score	E value	N
gb|AA074703|AA074703	zm76g07.r1 Stratagene neuroepitheli...	2071	4.0e-167	1
gb|AA068680|AA068680	mm61a05.r1 Stratagene mouse embryon...	1270	4.4e-145	4
gb|AA134788|AA134788	zm81g02.r1 Stratagene neuroepitheli...	946	1.3e-144	5
gb|AA399110|AA399110	zt89e06.s1 Soares testis NHT Homo s...	520	8.7e-119	6
gb|N39612|N39612	yy51g06.s1 Homo sapiens cDNA clone ...	582	9.6e-110	7
gb|AA282175|AA282175	zt02d03.s1 NCI_CGAP_GCB1 Homo sapie...	771	9.4e-80	3
gb|W81382|W81382	zd86f01.s1 Soares fetal heart NbHH1...	329	1.6e-75	6
gb|AA544946|AA544946	vk38e02.r1 Soares mouse mammary gla...	644	9.6e-63	2
gb|W35374|W35374	zc07h03.s1 Soares parathyroid tumor...	294	4.5e-42	4
gb|W57106|W57106	md57c12.r1 Soares mouse embryo NbME...	394	1.9e-30	2
gb|AA244877|AA244877	mx25a04.r1 Soares mouse NML Mus mus...	162	2.1e-27	4
gb|AA017857|AA017857	mh44d10.r1 Soares mouse placenta 4N...	230	3.7e-23	3
gb|AA531006|AA531006	nj07b11.s1 NCI_CGAP_Pr22 Homo sapie...	139	2.3e-19	3
gb|H32171|H32171	EST107015 Rattus sp. cDNA 5' end.	207	2.6e-10	2
gb|W79710|W79710	zd86f01.r1 Soares fetal heart NbHH1...	157	0.0073	1

TABLE 6

***mcg7*-specific oligonucleotides**

name	sequence (5' to 3')	SEQ ID NOs.
M1044R	GGA CAA AGT GTG TGA TGA ACC	SEQ ID NO:25
MCG7-GEF-REV2	CTC ATC CTC CGT CTG ATA CTG	SEQ ID NO:26
M7R	GTA GAT GTG GAT CAG CTT GG	SEQ ID NO:27
MCG7 CA FOR	AGG TGG AGA ATG GTC AAGG	SEQ ID NO:28
MCG7-GEF-REV	GTC ATA GTC TGT CTC CTA CT	SEQ ID NO:29
MCG7 GEF FOR	ACA TAG ACA GCG TGC CTA CC	SEQ ID NO:30
MCG7-PKC-REV	TAC AAC CTT AGG GAC ACC AG	SEQ ID NO:31
MCG7-PKC-FOR	TGC TGA GCC TGC TCA CGG TG	SEQ ID NO:32
T09103F	CAA GTG AAC AGC ACG TCC	SEQ ID NO:33
M7F	GAC TAT CTC AAG GAC CAG CTG	SEQ ID NO:34
MCG7UF	GGT TCG GTC CGA GCC CGG	SEQ ID NO:35
SGCADRV2	GGA GCG ATA CTC CAA GTA GGT	SEQ ID NO:36

TABLE 7

***mcg18*-SPECIFIC OLIGONUCLEOTIDES**

	name	sequence 5' to 3'
5	HVESTF	AGC GGG CCA GGC CCC TTC [SEQ ID NO:37]
	HV195F	CAT CCT GGT CCA ATG CGC TC [SEQ ID NO:38]
	HV387F2	GCA CTG AGG AAG TTA AAC GAG C [SEQ ID NO:39]
	HV408R	GCT CGT TTA ACT TCC TCA GTG C [SEQ ID NO:40]
	EXON1REV	GCT CAG CTC CAC AAA GCG GCT [SEQ ID NO:41]
10	HVEST426F	ACC AGC TCC GCT CAG GTA G [SEQ ID NO:42]
	HVEST623R	TCC AGG AGC TGT GTG TTT GG [SEQ ID NO:43]
	SGVESTF3	CCA GTT TCA CAG CGT GAG G [SEQ ID NO:44]
	HVEST631R	CAG CAT GAG GAG GAG GCA G [SEQ ID NO:45]

TABLE 8

EST CLONE SEQUENCES USED TO GENERATE HUMAN AND MOUSE *mcg18* cDNA SEQUENCE COMPOSITES

EST clone number	organism	GenBank accession number
1g2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

BIBLIOGRAPHY

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol. 215:* 403-410.

2. Church, G., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA 18:* 1991-1995.

3. Porfiri *et al. J. Biol. Chem. 269:* 22672-22677 (1994).

4. Sambrook, J., Frtisch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA.

5. Smith, D.B., and Johnson, K.S. (1988) *Gene 67:* 31-40.

6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) *Nucleic Acids Res. 22:* 4673-4680.

7. Beranger, F., Paterson, H., Powers, S., de Gunzburg, J. and Hancock, J.F. (1994) *Molecular and Cellular Biology 14:* 744-758.

8. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) *Genome Res. 6:* 124-131.

9. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun. 220:* 922-928.